DOCUMENT RESUME

ED 297 720 IR 013 425

AUTHOR Collis, Betty; Moonen, Jef
TITLE Designing an External Evaluation of a Large-Scale Software Development Project.
PUB DATE 7 Apr 88
NOTE 14p.; Paper presented at the International Association for Computing in Education Conference (New Orleans, LA, April 5-9, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Software; Conferences; Developed Nations; Elementary Secondary Education; *Evaluation Criteria; Evaluation Methods; Foreign Countries; *Formative Evaluation; *National Programs; *Needs Assessment; Program Development; *Program Evaluation
IDENTIFIERS *POCO Project (Netherlands); *Software Development; Stakes Model

ABSTRACT

This paper describes the design and implementation of the evaluation of the POCO Project, a large-scale national software project in the Netherlands which incorporates the perspective of an evaluator throughout the entire span of the project, and uses the experiences gained from it to suggest an evaluation procedure that could be applied to other educational software projects. The evaluation procedure for the first cycle of software development and distribution, which extends from September 1987 to January 1989, involves five components: (1) evaluating the intentions of the project based on the perspectives of the Minister of Education, which is funding the project, and of key members of the management team; (2) documentation of actual program activity; (3) reassessment of intended processes and outcomes; (4) evaluation of outcomes and project status; and (5) recommendations for program adaptation, based on three activities--creation of a preliminary priorities list, a working conference of experts, and preparation of a "white paper." (15 references) (EW)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED297720

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.



# Designing an External Evaluation of a Large-Scale Software Development Project

Betty Collis
Department of Psychological Foundations
Faculty of Education
University of Victoria
Victoria, British Columbia
Canada V8W 2Y2

Jef Moonen
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Paper presented at the International Association for Computing in Education Conference, New Orleans, April 7, 1988.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Betty Collis
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR013425

BEST COPY AVAILABLE

Abstract

Although it is frequently recommended that an evaluation component be part of a development program involving educational applications of computers and other information technologies, few software development projects incorporate the perspective of an evaluator throughout the entire span of the project. The POCO Project in The Netherlands is a large-scale national software development project whose first cycle of software development and distribution extends over the period September 1987 to January 1989. An external evaluator is involved with the project throughout this period. This paper describes the design and implementation of the evaluation of the POCO Project and uses the experiences gained from it to suggest an evaluation procedure that could be applied to other educational software development projects.

## Designing an External Evaluation of a Large-Scale Software Development Project

Program evaluation is intended to provide valid and useful information to audiences concerned with the effective operation or future of a program. There are many critical decisions that must be made about the purpose, design, and implementation of an evaluation study before this sort of valid and useful information can be systematically collected and communicated to the intended audiences. The identification of some of these decisions and subsequent illustration of the decisions in the context of an actual large-scale software development project can be of value to those involved in decision-making positions in other projects pertaining to the development of educational software and accompanying support materials. This paper will briefly describe a national software development project in The Netherlands, outline the intentions of the project management team in commissioning an evaluation of the project, indicate some of the critical decisions in designing and implementing the evaluation, and discuss some ways in which the evaluation has been of value to the project. In addition, the paper will provide recommendations for similar evaluation studies for other projects involving new information technologies.

### The POCO Project

The program that is the object of the evaluation described in this paper is the POCO Project, announced by the Dutch Ministry of Education and Science in March 1987 and formally approved in May 1987 through a "policy note" (ECC, 1987). "POCO" is taken from the Dutch name of the project, "Programmatuur Ontwikkeling voor Computers in het Onderwijs"; in English, "Program Development for Computers in Schools." The major goal of the project is the production of courseware that can be directly used in existing curricula. Courseware is defined as teaching/learning materials consisting of computer software and accompanying support materials.

The Ministry specified that the development projects undertaken by the POCO Project must focus on materials that can be utilized by teachers in a meaningful way and with such frequency during their regular teaching activities that teachers will come to perceive using such packages to be an effective and efficient response to an educational need. "Frequent use of appropriate

courseware" is seen as THE major component in the process by which teachers become familiar with the use of computers.

The POCO Project will involve the implementation of four processes, each essential in the production of courseware:

- choosing priorities
- formulating product descriptions
- managing technical production
- distributing the courseware.

The target groups of this project are primary, general secondary, and lower/middle vocational education. The project consists of two cycles: September 1987 to January 1989, and January 1989 to September 1991. After the first cycle, the Dutch Ministry of Education and Science will decide upon the execution of the second cycle, based upon an evaluation of the first cycle. The total budget for the project is 24 million Dutch guilders. The project will be managed by the "Educational Computing Consortium - ECC," which is the privatized successor of the Centre for Education and Information Technology (COI), University of Twente, Enschede, The Netherlands.

## Purpose of the Evaluation

Defining the contractor's purpose in commissioning an evaluation is a critical step in the planning for any evaluation project (Collis, 1987a; Stufflebeam & Webster, 1980) and one of the particular factors that distinguish evaluation research from other types of research. In evaluation research the contractor rather than the researcher originates the research questions, at least in a general way, and delimits the parameters of the design of the study.

Evaluations may be commissioned in order to provide information for ongoing readjustments of program activities and goals, for funding decision relative to the continuation of a program, or they may be politically motivated. Frequently they may be issue oriented or involve the assessment of competing programs. With respect to the POCO Project, the motivation for including an evaluator as part of the project team from the beginning of the project was to provide information for ongoing readjustments of program activities and continuous quality control and assessment of the goals of the project.

## Critical Decisions in Designing an Evaluation Study

The purpose of the evaluation motivates the choice of a design for the study. A major decision that must be made in choosing a design model relates to the degree to which the contractors wish the evaluation to operate at a "goal-free" level (Scriven, 1973), or restrict its focus to prespecified project goals. Stake (1977) calls the latter "preordinate" evaluation and contrasts it to a more emergent or "responsive" form of evaluation design.

"Preordinate" designs require the prior specification of desired program outcomes. There must be some predetermined standard for program success against which the program outcomes are measured, and objective instruments or standardized measures are frequently employed in this measurement process. "Responsive" evaluation, in contrast, orients more directly to the program activities or "transactions" than to the program outcomes, and not only allows for but expects that different participants in a program will have different viewpoints concerning the success of a program, as well as of the appropriateness of the ongoing decisions within the program. A responsive approach involves the perspective that programs may evolve as they operate so that original goals and strategies are adapted or even abandoned as they are subjected to the ongoing responses of individuals involved in the program (Stake, 1977).

Both approaches have characteristic strengths and limitations, particularly with regard to the breadth and reproducibility of data collection (Schermerhorn & Williams, 1979). In responsive evaluations, the evaluation report often "is personal and vicariously conveys feelings as to what it is like to participate in the programme experience" (p. 55); however, the opportunity for evaluator bias is obviously heightened. In preordinate evaluations, the evaluation report is based as much as possible on objective, verifiable data and the evaluator intends to intrude on the system or data as little as possible. A responsive approach seems to be more appropriate when the focus of the evaluation is a group experience, such as a training course, a conference, or ongoing participation in a project within a workplace or school setting. A preordinate approach seems advised when a project has well articulated, measurable goals.

Most large-scale programs, such as the POCO Project in The Netherlands, involve both these types of components, in that there are many situations where both process and product are of interest. Stake has developed an evaluation model which reflects

both the process and the product dimensions (Stake, 1973) and which can be an appropriate design for the evaluation of software development and distribution projects like POCO (Shapiro, 1985).

Stake's model can be modified to involve three major components (Moonen, 1987). The first relates to a clarification of program intentions with regard to both the expected outcomes of a program and the activities that are planned to bring about those outcomes. The logical relationship between intended activities and intended outcomes is a component of this portion of the evaluation. A second component of the evaluation involves the observation of actual program activities and outcomes and yields an assessment of the congruence between what was intended and what actually occurred. Deviations from intentions are synthesized together with actual program outcomes in order to suggest a new set of modified intentions for both process and product that better reflect the realities of the project as it evolved. The recommendation of a new set of intentions for subsequent cycles of a program is the third component of the evaluation activity. An advantage of this model is that it facilitates the revision process if the intended outcomes of the program are not achieved to the degree expected. This model helps to distinguish underachievement due to "theory failure" from that due to "program operation breakdown" (Suchman, 1976), or "program slowdown." This distinction has critical implications for subsequent recommendations for program modification.

## Application of the Modified Stake's Model to the POCO Project

This adaptation of Stake's model was chosen as appropriate for the evaluation of the POCO Project. The goals of the POCO Project are ambitious: not only to produce and distribute relevant courseware that will be in "frequent use" by teachers in the primary, secondary, and vocational sectors of the Dutch educational system; but also to promote and, potentially, to market POCO products and expertise outside The Netherlands.

If the software goals, and subsidiary goals judged to be instrumental to the overall attainment of the project goals, are not being met as planned, it will be important to distinguish between theory failure and program slowdown. This will be especially pertinent at the completion of Cycle 1, as decisions will have to be made about continuation of the project into Cycle 2 and, if continuation occurs, about adaptations to both process and product expectations in Cycle 2. Theory failure would suggest the original directives for the POCO Project require modification or the original expectations were unrealistic because

of some number of situational variables. If this can be documented, the expectations of the funder of the project, the Minister of Education in The Netherlands, may have to be modified if program success is to occur. Program failure, in contrast, would not call for this type of global reconceptualization of the overall POCO goals but instead would suggest small and large adjustments in various component parts of the ongoing POCO activities.

Within this framework, a five-component evaluation design has been developed for the POCO Project (Collis & Bergers, 1987). Each of the components is structured around a set of critical research questions. The five components are described in the following subsections.

## Evaluating the Intentions of the POCO Project

Component 1 of the evaluation begins with the delineation of the intentions of the project as of September 1987 based on the perspectives of the Minister of Education, who funds the project, and of key members of the management team. Intentions relate to both the anticipated activites of the project and the expected outcomes and status of the project as of January 1989. Evaluation questions based on these intentions relate to the degree of consensus that exists among key people involved in the project with regard to the stated and unstated motivations for the project, the evolution of the "priority" list of software products for actual development, and the development and field testing/revision components of the Cycle 1 activities. Of particular importance is the perception of who is responsible for which decisions in each of these areas, particularly when this perception varies for different key figures involved in the project.

## Observation of Actual Program Activity

The second component of the evaluation involves the documentation of what actually occurs during the execution of the project over the sixteen months of Cycle 1. Special consideration will be given to instances where program activity as it occurred did not match what was expected.

## Reassessment of Intended Process and Outcomes

Based on the ongoing assessment of what actually occurs during Cycle 1, the third component of the evaluation project involves the prediction of the likely impact on the intended

outcomes of the project of the particular program activities or planning as they actually transpire. Were alterations in expected procedures sufficiently significant so that it is no longer likely to expect the intended outcomes to occur? What sort of modifications in expectations should be made?

Evaluation of Outcomes and Project Status

The fourth component of the evaluation will examine the actual outcomes of the project as of January 1, 1989 and compare these outcomes with those that were originally expected for the project at that point in time. If a discrepancy between intentions and actuality occurs, the evidence accumulated throughout the evaluation will be used to distinguish between theory failure and program operation slowdown.

Recommendations for Program Adaptation

The final component of the evaluation study will be a set of recommendations pertinent to the second cycle of the POCO Project based on the experiences gained during Cycle 1.

## Application of Stake's Model to the Choosing of Priorities Within the POCO Project

This adaptation of Stake's model can be applied to the evaluation of an individual component of a project at the same time that it is being applied to the overall project. Within the POCO Project, this occurred by evaluating the first major phase of the project--the "choosing priorities" phase--and at the same time evaluating a "working conference" of the project that was held in Enschede, The Netherlands, on September 21-26, 1987 (Collis, 1987b). The goal of the first phase was to establish a priority list of software to be produced within the first cycle of the project. The working conference was the second activity (in a series of three) within this phase of the POCO Project.

The first activity within this phase had resulted in a preliminary priority list, which was prepared following discussions from a conference on September 9-11, 1987, involving Dutch experts from the different educational sectors related to the POCO Project. During the third activity of this series, so-called "white papers" were written, in which components of the established priority list were discussed from curricular and organizational-logistics standpoints (Nagtegaal, 1987). For primary, general secondary, and vocational education respectively,

5, 15, and 16 white papers were written. These white papers were presented to the Minister of Education and Science on December 1, 1987. As agreed beforehand, he selected respectively 2, 8, and 8 of them by January 15, 1988. These white papers form the basis for the second phase of the project: the development process.

## Evaluation of the Working Conference

The working conference had both measurable and unmeasurable goals. The measurable goals involved the identification of specific, already available software packages that might be useful to incorporate with the materials being developed by the project, and the suggestion of revisions to the priority list for the content and scope of software to be developed during Cycle 1 of the project. The unmeasurable goals relate to the development of internal cohesiveness and commitment to the POCO Project among representatives of the Dutch educational community involved in the project, and to the development of a positive reputation for the project both inside and outside The Netherlands (ECC, 1987).

Data were collected through questionnaires and interviews but were primarily obtained from the observations of the evaluator based on her experience of "what it was like to participate in the programme experience" (Schermerhorn & Williams, 1979, p. 55). All these sources of data helped address the preordinate aspects of the evaluation: To what extent did the working conference meet its goals relating to software selection and revision of the priority list? They also were employed to evaluate the success of the working conference with respect to the "unmeasurable" goals of developing commitment to the project.

The general conclusions of the evaluation study were that the working conference was an effective way to nurture the goals of the POCO Project with respect to: (a) strengthening the perception in the Dutch educational community and abroad that the project will be productive, professionally managed, and will make a significant contribution to educational computer usage; and (b) suggesting clarifications for the priority list, particularly with respect to tool-type software which can be used across educational sectors and in an interdisciplinary manner. Also, the general plan for the working conference--with foreign specialists, morning presentations, afternoon software demonstrations, and evening discussions--was judged to be an effective approach and was recommended again for a Cycle 2 working conference (Collis, 1987b).

From a responsive perspective, the evaluation study also focused extensively on the ongoing activities of the working conference. On a day-to-day basis each particular activity was evaluated with respect to its ultimate contribution to the overall goals of the working conference and of the POCO Project generally, and recommendations were made on a daily basis when some aspects of project activity appeared to not work as intended. This type of analysis is highly dependent on the observations obtained within the responsive framework of the evaluation model; it generated a list of specific recommendations for modifications and adaptations of intended activities for a similar working conference to precede Cycle 2 (Collis, 1987b), but also provided ongoing feedback to the management team resulting in daily modification to the working conference as it proceeded.

## Evaluation of the First Phase: The Choosing of the Priorities

During the three months following the working conference, the POCO team worked on the development of the educational rationale for the priorities being selected as potential focuses for subsequent courseware production. Also the team addressed issues relating to its own infrastructure and functioning and made frequent contacts with particularly important members of the Dutch educational community--the Ministry, "Cluster I" personnel, and the educational publishers. The role of the evaluator during this period was to observe and collect information and to submit two documents (Collis, 1987c, 1988). The function of each of these documents was proactive and consultative. The first document, presented to the management team on December 4, 1987, identified a list of critical decisions that need to be addressed by the team, suggested alternative responses to those decisions, identified a time line during which the decisions must be made, and predicted consequences of various responses to the decisions. This document was the focus of discussion during an intensive, two-day team meeting, December 14-15, 1987. The second evaluation document, presented January 2, 1988, reconsidered the critical issues identified in the December 4 document, in light of progress made and renominated critical issues and possible responses. In addition, the document served a different type of proactive role by including a suggested plan for addressing the public relations aspect of the project.

## Conclusions

From the perspective of the management team of the POCO Project, the contributions of the ongoing evaluation component within the project relate both to its responsive aspects, allowing

an ongoing quality control assessment relating to the activities of the project, and to its preordinate aspects, used to assess and adjust the theoretical and operational goals of the project during its operation. It is hoped that, by this approach, traditional conflicts that arise between executors and evaluators relating to the description of the goals, responsibilities, and the exactness of data pertinent to an already finished project can be avoided. The utilization of an ongoing evaluation should, on the contrary, establish an atmosphere through which one accepts that the execution of a project involves making mistakes, most of them inevitable due to particular internal and external circumstances. At the same time the ongoing evaluation provides the explicit opportunity to identify these circumstances, to learn from them, and to avoid making those mistakes again in the next phase or cycle of the project.

The experiences of the POCO evaluation can also be used to support a recommendation to other managers of comparable projects that they include an ongoing evaluation study as an essential part of the projects. In such a way the evaluation study will be likely to have the most value to the project itself. Too often, evaluation studies, conducted afterwards, have only a minor influence because (a) critical remarks occur too late to be incorporated into ongoing project activity, and (b) the results of the study will not be likely to be used for new projects because each project creats its own circumstances and contextual conditions. The evaluation design used in this project allows for implementation adjustment as well as goal respecification, and as such is recommended for other large-scale software development and implementation projects.

References

Collis, B. (1987a, June). Evaluating a school or district computer program. Paper presented at the National Educational Computing Conference, Philadelphia, PA.

Collis, B. (1987b). Evaluation study of the POCO Project "Working Conference". Report presented to the Educational Computing Consortium (ECC), Twente University, Enschede, The Netherlands.

Collis, B. (1987c). POCO Evaluation Report 2: The Priority Determination Phase. Report presented to the Educational Computing Consortium, Twente University, Enschede, The Netherlands.

Collis, B. (1988). POCO Evaluation Report 3: Planning for 1988. Report presented to the Educational Computing Consortium, Twente University, Enschede, The Netherlands.

Collis, B., & Bergers, L. (1987). Evaluation design: POCO Project Cycle 1. Report submitted to the Educational Computing Consortium, Twente University, Enschede, The Netherlands.

ECC. (1987). The POCO Project. POCO-Bulletin, Number 1. Educational Computing Consortium, Enschede, The Netherlands.

Moonen, J. (1987, December). Applying a general evaluation model to a seminar on evaluation of national policies for educational computing. Presented to EDITE 87, Evaluation and Dissemination of Information Technologies in Education, Luso, Portugal.

Nagtegaal, C. (1987). Prioriteiten POCO: Advies en verant woording. Enschede, The Netherlands: Educational Computing Consortium.

Schermerhorn, G. R., & Williams, R. G. (1979). An empirical comparison of responsive and preordinate approaches to program evaluation. Educational Evaluation and Policy Analysis, 1(3), 55-60.

Scriven, M. (1973). The methodology of evaluation. In B. R. Worthen & J. R. Sanders (Eds.), Educational evaluation: Theory and practice (pp. 60-104). Belmont, CA: Wadsworth Publishing Company.

Shapiro, J. Z. (1985). Evaluation of a worksite programme in health science and medicine: An application of Stake's model of contingency and congruence. Educational Evaluation and Policy Analysis, 7(1), 47-56.

Stake, R. E. (1973). The countenance of educational evaluation. In B. R. Worthen & J. R. Sanders (Eds.), Educational evaluation: Theory and practice (pp. 106-125). Belmont, CA: Wadsworth Publishing Company.

Stake, R. E. (1977). To evaluate an arts programme. Unpublished report, University of Illinois at Urbana-Champaign.

Stufflebeam, D. L., & Webster, W. J. (1980). An analysis of alternative approaches to evaluation. Educational Evaluation and Policy Analysis, 2(3), 5-20.

Suchman, E. (1976). Evaluation research. New York: Russell Sage Foundation Press.